DOCUMENT RESUME

ED 057 088 TM 000 931

AUTHOR Clifford, Margaret M.
TITLE Decision-Making Rationale for Educational Testing.
INSTITUTION Iowa Univ., Iowa City. Coll. of Education.
PUB DATE [Oct 71]
NOTE 13p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Achievement; *Achievement Tests; Behavioral Objectives; *Decision Making; *Educational Testing; *Individual Power; Measurement Techniques; Student Motivation; *Student Opinion; Student Participation; Student Role; Test Construction; Test Interpretation; Test Results
IDENTIFIERS *Decision making Rationale for Educational Testing; DRET

ABSTRACT

Decision-making Rationale for Educational Testing (DRET) is a proposal intended to reduce the misuse of achievement tests. It assumes 1) measurement is intended to facilitate decision-making, 2) the choice of an instrument, the identification of examinees and the use of test results is determined by the decision for which the measurement is to be taken, and 3) effective educational measurement is a function of the nature of the decision, the examinee's option to measure, and his expectation of the measurement. DRET specifies that a decision issue must be clearly stated and validated before initiating measurement, and that measurement in excess of that which facilitates the decision is prohibited. This proposal gives the student a major role in determining measurement activities and might be summarized in the motto, "Test at Student Request." (Author)

U S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
IS DOCUMENT HAS BEEN REPRO-
UCED EXACTLY AS RECEIVED FROM
E PERSON OR ORGANIZATION ORIG-
ATING IT. POINTS OF VIEW OR OPIN-
NS STATED DO NOT NECESSARILY
EPRESENT OFFICIAL OFFICE OF EDU-
ATION POSITION OR POLICY

ED057088

Decision-Making Rationale for Educational Testing

Consider the possibility of a well-enforced policy which stipulates that achievement tests may be administered only at the request of a student; written justification must be given for every measurement activity conducted in the school; and confidentiality of test results must be insured. This would represent a major change in current measurement programs and possibly a solution to many related problems.

There is discontent with educational measurement as evidenced by the numerous attacks which have been made on standardized tests. It is said that they discriminate unfairly, are culturally biased, are products of money-making organizations, are an invasion of privacy, are a cause of unwarranted frustration and anxiety, and are outdated and irrelevant (Gross, 1962; Hoffman, 1961; Yourman, 1970). Nongraded programs have been introduced in an attempt to de-emphasize levels of achievement (Rhoades, 1966; Perkins, 1961; Goodland & Anderson, 1959). The pass-fail option, and the use of a Satisfactory-Unsatisfactory scale were initiated to minimize distinctions in performance (Sgan, 1969). Parent-teacher conferences as opposed to the quarterly or six-week report cards, are another attempt at eliminating the quantifying and categorizing of students. A "good teacher" tends to be equated with one who avoids tests, de-emphasizes grades, and refrains from using norms (Rogers, 1961).

Is this a backlash against our scientific passion for precision? Or is measurement the most vulnerable spot in education at a time when protest appears to be in style? Perhaps if we step away from the last test we took, the most recent exam we prepared, or the

TM 000 931

report card we just signed, it will be possible to get an objective view of measurement, and identify more accurately the nature of the problem of educational measurement.

Measurement is an acceptable and respectable activity in the physical and domestic sciences. The refinement of measurement has usually been recognized as a mark of progress, and the limitations of an instrument frequently supplied the motivation for such progress. Impatience with imprecision resulted in the perfection of instruments more frequently than in their elimination. Dissatisfaction with the sun dial, for example, did not lead to a disregard for the time of day, but rather to the invention of precision time pieces.

Educational measurement, however, has enjoyed comparatively little prestige or progress. The limitations of testing programs are vigorously discussed, but there has been a shortage of constructive criticism. Perhaps we lack the patience that marked progress in measurement requires; perhaps we doubt the need for educational measurement. But if we are intent on improving education, we must resolve our problems with measurement. For improvement or progress without some form of measurement is questionable.

The purpose of this article is to propose a Decision Rationale for Educational Testing (DRET) which would provide a guide for developing effective measurement programs in education. The proposal is primarily concerned with the measurement of academic achievement rather than psychological or attitudinal measurement. DRET implies that the use of a measuring device must be accounted for in terms of a decision to be made on the basis of such measurement. In addition to this decision factor, there is a factor of expectation and a

factor of option which are believed to be significantly related to the effects of educational measurement. After briefly exemplifying the role of these three factors (i.e., decision, expectation, and option) in noneducational measurement, I will discuss their function in educational measurement.

Ordinarily an individual has a definite reason for using a tape measurer, for checking the outdoor thermometer, or for computing mileage on a trip. It is the reason which justifies one's reading, recording, adding, subtracting or comparing of numbers. A person records mileage, not only to satisfy curiosity, but to decide what figures to write on an expense account, or which of two routes is the shorter, or whether a recent tune-up improved his mileage. A pending decision motivates the act of measuring.

Accompanying the need for a decision there is usually a factor of expectation which affects the satisfaction derived from measurement. It is more enjoyable for an individual to test his tennis skill when he expects improved performance than when he is simply asked to serve as a practice partner. The task of meticulously recording experimental data becomes less boring and possibly exciting when an individual anticipates significant effects. Expectation reflects a personal interest that makes measurement meaningful.

A third characteristic of most measuring activities is that of option. We are free to take or not take a driver's test; to regard or disregard a speedometer; to step on or over a scale; to measure or guess at the amount of paint needed for the house. There are limits, restrictions, and consequences which affect our choice in

most cases, but an individual is generally free to measure or avoid measurement.

What decisions, expectations, and options are involved in educational measurement? At present, the answer to this question leaves much to be desired. I suggest that a major reason for educational testing is administrative routine: It is time to give report cards and evidence is needed in case of student or parent objections. It is an odd numbered year and IQ tests are given alternately with achievement measures. The Title I Project Report requires pre- and posttest comparisons. The school board wants the annual results of the city-wide testing program. Funds for a physical fitness program have to be justified.

What decisions are at stake? How important are they? For whom are they important? A teacher can decide which of her pupils will get A's and who will be given the B's, C's, etc. The test results from the Title I Program may have an effect on future funding. The results of the city-wide testing may lead to a decision regarding which of several experimental programs should be given preference for the coming year. But, these types of decisions are administrative. Measurement is imposed upon students so teachers and administrators can solve their problems or satisfy their responsibilities. The decision power is not in the hands of the individual most immediately responsible for the measurement. The nature of the decision is often of little or no interest to the student. The decision process seldom involves him beyond the point of data gathering.

The option factor is virtually nonfunctional in the measurement activities of our schools. A student may arrange to be ill during

an exam period, but aside from such a manuever he has little or no choice about when, how, or on what content he is to be measured. He has no right to say, "I'll skip the standardized tests this year;" nor does he enjoy the freedom of saying, "I'd like to measure my science and math achievement, for I'm considering a course in physics."

The factor of expectation and its effect on educational measurement is difficult to assess. It seems reasonable, however, that expectation will be more specific on freely selected tasks than on imposed tasks. If expectation functions as a goal-setting motivator (De Cecco, 1969), intensifying student expectation can be expected to improve test performance, and should, therefore, be one goal in designing a testing program.

In summary, measurement programs as they currently exist in our schools are not for purposes which the student defines; he is not a decision-maker; there is little reason to believe he has strongly motivated expectations; and he has probably never enjoyed less option. Precision instruments would probably be of little value under these circumstances, and the most crudely designed could not be held accountable for the resulting discontent. Thus, a resolution of our educational measurement dilemna requires more than a technical discussion of instruments. It requires a re-evaluation and reformulation of our rationale and a corresponding modification of our testing policies.

DRET is not an elaborately delineated program ready for adoption, but rather a broad framework on which any number of meaningful measurement programs can be developed. It entails three assumptions

and two ground rules. It assumes 1) measurement is intended to facilitate decision making; 2) the choice of an instrument, the identification of examinees, and the interpretation and use of results is determined by the decision for which the measurement is to be taken; and 3) effective educational measurement is a function of the nature of the decision, the examinee's option to measure, and his expectation of the measurement. The ground rules are:

1. A decision issue must be clearly stated and validated before measurement is initiated.
2. Measurement in excess of that which directly facilitates the validated decision-making activity is prohibited.

A look at current educational practices in view of these rules clearly emphasizes the excessiveness of our measurement activities: It is difficult to imagine a decision which would require the routine testing of every student at a given level over all major areas of content. For a system-wide curriculum decision, one would only need to measure a random sample of students over just that content relevant to the decision issue. A decision concerning the individual student would require testing the one individual for whom a decision is to be made. It is inconceivable that a school system makes an annual decision about every 4th, 6th, and 8th grader on every major subject during the month of October or November!

To envision a DRET-based testing program, suppose that a school system had a collection of measurement instruments--a test library. This might be a simple compilation in a filing cabinet or a sophisticated computerized bank. The library includes achievement, aptitude, and ability instruments. At the checkout desk or terminal there is a

sign that reads, "TEST AVAILABLE AT EXAMINEE REQUEST ONLY!" Aside from a few rare exceptions, no instrument is "checked out" unless it is requested by the examinee for a specific and justifiable purpose implying some decision. Only that part of the instrument related to the decision issue is loaned and provisions for its supervised use are made. The test results are personal information to be revealed or shared only at the discretion of the examinee (or parents in the case of a very young child).

Under such a system the examinee's option to measure would be guaranteed. Requiring the specification of the decision for which measurement is to be used, would help insure test validity and maximize student effort. The expectation factor would undoubtedly be operative, since the examinee not only can but must determine when, why, and on what content he will be measured. Testing could become a measurement activity which is under the immediate control of the examinee--assuming the examinee can be motivated to measure his academic achievements!

The motivation required to operationalize a "test at student request" program could be a challenge. It is obvious that students will have to be given incentives--acceptable reasons--for taking tests. But, then, an improved rationale is desired not only by students, but teachers and administrators as well. I propose that an emphasis on pretesting as opposed to posttesting would facilitate a DRET-based measurement program.

Few academic endeavors are valuable in themselves. They usually acquire importance only in view of future activities. The mastery of shorthand I is rather inconsequential unless shorthand II or a

similar activity requiring the application of shorthand I is desired. On the other hand, the course content, the teaching method, and the class composition usually assumes a given level of achievement which supposedly helps insure success. Thus emphasizing conditions for course enrollment (i.e., stating requisites and prerequisites) would have greater relevance than emphasizing conditions for successful course completion. Test taking would become a means by which students would assure themselves and the instructor that they were prepared for a given learning experience.

Furthermore, if prerequisites are expressed in terms of achievement levels rather than completed courses or specific experiences, the means to achievement can be left flexible. The student can simply be held accountable for providing evidence of achievement. A shift of emphasis from post-measurement to pre-measurement would offer improved guidance for course selection as well as protection against failure. The use of requisites and prerequisites might also be used to provide incentives for attemding to less desirable learning activities. For example, the student whose primary interest is shop and whose least desirable subject is reading might discover that the requirements for the first shop class are:

1. To demonstrate at least a 4th grade level of reading ability and a 3rd grade level of math ability.
2. To carry simultaneously a minimum of two additional subjects in which satisfactory performance is demonstrated.

The requirements for his second shop class may be:

1. To have demonstrated average proficiency on the skills taught in the preceding shop course.

2. To have successfully completed one course in art design.
3. To have demonstrated at least a 5th grade level of reading ability.
4. To carry simultaneously a minimum of two additional subjects in which satisfactory performance is demonstrated.

By the time the student reaches the 4th or 5th shop course the requirements might include "having an academic program that can terminate in a diploma or certification within 12 months." Students (with varying degrees of counseling) would decide upon a program or course, and attempt to demonstrate their qualifications (typically by providing test scores). Substandard performance (i.e., below prerequisite level) on an instrument would simply suggest to an examinee that additional knowledge or skill is needed prior to attempting his selected course. The individual who "measured up" to a course would request that his score be forwarded to the appropriate teacher, and thus earn admission. Students would test into courses rather than out of them.

There always are and, undoubtedly, always will be students who lack motivation; who appear to have no course interests and therefore would have no reason for measuring their ability or achievements. A DRET approach, however, is more likely to decrease than increase membership in this fold. Increasing the student's decision power and his options in program development and measurement can be expected to reduce student apathy.

This brief presentation of a DRET-based measurement program raises many questions: Who determines the prerequisites and the corresponding measurement instruments? What kind of norms or standards are most meaningful and under what circumstances? How and by whom are scores

interpreted? What happens to the student who is poorly motivated for learning, has no desire to measure his skill, has no prerequisites he cares to meet? The cooperation and collaboration of educators, test constructors, and measurement experts can undoubtedly arrive at a variety of acceptable answers to these and related questions. The major issue is whether the proposal justifies an attempt to overcome the problems it may entail.

Five practical questions related to such a program are briefly discussed below:

1. How would a DRET program affect test construction? There would be an increased demand for the services of test constructors--a need for item pools from which teachers could compile reliable and valid measures, as well as intact tests and subtests. Test constructors would probably be asked to make and standardize instruments for individual texts, units, modules and courses.
2. How is standardization achieved when students are no longer obliged to take tests? One might simply explain to students the importance of well-developed measuring instruments and ask them to assist in test construction or revision. On the other hand, we might standardize instruments on the basis of those students who chose to use the instrument during the year.
3. What happens to the 6 or 9 week tests and the report card routine? A multitude of communication techniques could be developed for reporting educational progress. For example, each time a student begins a unit, course, or program, parents could be notified that, "On the basis of having successfully met the following prerequisites, (List them), (Name of Student) has been admitted to (Unit, Course,

Program)." In addition to these communications, a conference method might be used to discuss a student's progress.

4. Would teacher-made tests, typically given on a daily or weekly basis, be outmoded? Not at all. Such tests would be made available for students who desire a measurement on the specified content. A major function of such tests or quizzes would be to give the student an expectation of his knowledge and ability. This, in turn, would serve as a guide in his eventual selection of a comprehensive measurement required for a more advanced course.

5. Could tests no longer be used to measure the attainment of teacher-made goals (e.g., behavioral objective)? Teachers may find that "testing at student request" is a more meaningful way of assessing teaching effectiveness than is an imposed testing program. The DRET approach would require a teacher to make his goals become the goals of the student. Only to the extent that the teacher could successfully accomplish this, could he expect students to share concern for the measurement of goal attainment. The proportion of students who chose not to take a test or performed unsatisfactorily might suggest the extent to which goal attainment was unsuccessful. Test records could play a more significant role than ever in the assessment of goal attainment.

In summary, measurement in itself is a harmless activity which suggests a concern for progress or change. The problems we experience in educational testing probably reflect a need to challenge our measurement rationale more than our measurement instruments. We may be in need of better tests, but we do not deserve them until we find better things to do with them. Habit, tradition, and the accumulation

of reference materials are invalid excuses for testing programs. We must establish a rationale that will justify educational testing for the test constructor, the test administrator, and most of all, the test taker.

REFERENCES

De Cecco, J. P. The psychology of learning and instruction: Educational psychology. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1968.

Goodland, J. I. and Anderson, R. H. The nongraded elementary school, New York: Harcourt, Brace and Co., 1959.

Gross, M. L. The brain watchers. New York: Random House, 1962.

Hoffman, B. The tyranny of multiple-choice tests. Harper's Magazine, 1961, 222, 37-44.

Perkins, H. V. Nongraded programs: What progress, Educational Leadership, 1961, 19, 166-169.

Rhoades, W. M. Erasing grade lines. The Elementary School Journal, 1966, 67, 140-145.

Rogers, C. On becoming a person. Boston: Houghton Mifflin, 1961, 275-278.

Sgan, M. R. The first year of pass-fail at Brandeis University: A report. Journal of Higher Education, 1969, 40, 135-144.

Yourman, J. The case against group IQ testing. In Clarizio, H. F., Craig, R. C., & Mehrens, W. A. (eds.), Contemporary issues in educational psychology. Boston: Allyn and Bacon, Inc., 1970, 371-375.